

IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF PENNSYLVANIA

IN RE: ZOLOFT (SERTRALINE HYDROCHLORIDE) PRODUCTS LIABILITY LITIGATION	:	MDL No. 2342 12-md-2342
	:	
	:	
THIS DOCUMENT APPLIES TO:	:	
	:	
ALL ACTIONS	:	
	:	

MEMORANDUM OPINION

Rufe, J.

June 27, 2014

Plaintiffs in this multi-district litigation (MDL 2342) allege that the antidepressant Zoloft, when taken during pregnancy, caused birth defects in the children born to exposed mothers. The Plaintiffs' Steering Committee ("PSC") in MDL 2342 proposes to offer the testimony of various expert witnesses on the issue of general causation. These expert witnesses include Anick Bérard, a perinatal pharmacoepidemiologist, who holds a Ph.D. in Epidemiology and Biostatistics from McGill University, and who teaches at the Université de Montréal. Dr. Bérard has conducted research on the effect of drugs, including antidepressants, on human fetal development, and opines that Zoloft, when used at therapeutic dose levels during human pregnancy, is capable of causing a range of birth defects (i.e., is a teratogen).¹

Before the Court is the Motion to Exclude the Testimony of Dr. Bérard, filed by Defendants Pfizer Inc. and Greenstone LLC ("Defendants" or "Pfizer"). Pfizer does not challenge Dr. Bérard's academic qualifications, but argues that unreliable methods and principles were used to reach her conclusion that Zoloft may cause birth defects in the children of exposed

¹ Dr. Bérard does not opine as to whether Zoloft caused the particular malformations detected in any particular child (i.e. specific causation).

mothers. The Court has reviewed Dr. Bérard's report, as well as Defendants' rebuttal expert reports and the briefs of the parties, and held a *Daubert* hearing at which testimony and evidence were presented in support of each position.

I. STANDARD OF REVIEW

Federal Rule of Evidence 702 reads:

A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

- (a) the expert's scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
- (b) the testimony is based on sufficient facts or data;
- (c) the testimony is the product of reliable principles and methods; and
- (d) the expert has reliably applied the principles and methods to the facts of the case.

The Third Circuit has distilled this rule to two essential inquiries: 1) is the proffered expert qualified to express an expert opinion; and 2) is the expert opinion reliable?² With regard to Dr. Bérard, Pfizer challenges the reliability of the opinions.

Under the Third Circuit's framework, the focus of the Court's inquiry must be on the experts' methods, not their conclusions. Therefore, the fact that Plaintiffs' experts and Defendants' experts reach different conclusions does not factor into the Court's assessment of the reliability of their methods.³ The experts must use good grounds to reach their conclusions, but not necessarily the best grounds or unflawed methods.⁴

Here, the scientific question that Dr. Bérard has been asked to address is whether she believes that Zoloft may cause birth defects in children born to exposed mothers, to a reasonable

² *In re TMI Litig.*, 193 F.3d 613, 664 (3d Cir. 1999).

³ However, where the scientific community considers the evidence to be inconclusive, a difference of opinion may sometimes undermine the reliability of an expert's conclusion that there is a causal link, and may justify excluding that expert. *Magistrini v. One Hour Martinizing Dry Cleaning*, 180 F. Supp. 2d 584, 607 (D.N.J. 2002), *aff'd* 68 F. App'x 356 (3d Cir. 2003).

⁴ *Holbrook v. Lykes Bros. S.S. Co.*, 80 F.3d 777, 784 (3d Cir. 1996); *In re Paoli R.R. Yard PCB Litig.*, 35 F.3d 717, 745 (3d Cir. 1994).

degree of scientific certainty. To meet the *Daubert* standard, she must demonstrate that she has good grounds for her causation opinion (*i.e.*, the opinion is based on methods and procedures of science, not subjective belief) and a reasonable degree of scientific certainty regarding her causation opinion.⁵

Expert evidence must be relevant and reliable to be admissible. The Court must consider: 1) whether the expert's theory can be tested; 2) whether studies have been subject to peer review and publication; 3) the potential for error in a technique used; and 4) the degree to which a technique or theory (but not necessarily a conclusion) is generally accepted in the scientific community.⁶ The burden is on Plaintiffs to demonstrate that Dr. Bérard used reliable scientific methods to reach her opinions.

II. BACKGROUND

Zoloft is a prescription antidepressant, commonly used to treat depression, anxiety, and other mental health conditions. The active ingredient in Zoloft is sertraline. Zoloft is one of a class of drugs known as selective serotonin reuptake inhibitors (SSRIs). Serotonin is a neurotransmitter produced endogenously by humans and other animals. The SSRIs do not contain serotonin; rather, they alter the availability in the nervous system of the serotonin produced by the body. The FDA categorizes Zoloft as a Pregnancy Category C drug.⁷

⁵ See *Daubert v. Merrell Dow Pharmas., Inc.*, 509 U.S. 579, 590 (1993).

⁶ *Id.* at 593-94.

⁷ The FDA has established 5 categories to indicate the potential of a drug to cause birth defects if used during pregnancy. Category A means that there are adequate, well-controlled studies which have failed to demonstrate a risk to the fetus. Few drugs are in category A because controlled studies of medication use during pregnancy are ethically prohibited. Category B means animal studies show no risk, but there are no adequate and well-controlled studies of use by pregnant women. Category C means that animal reproduction studies have shown an adverse effect on the fetus, but there are no adequate and well-controlled studies in humans, and so pregnant women should weigh the potential benefits against the potential risks. Category D is used when there is positive evidence of human fetal risk based on adverse reaction data from investigational or marketing experience or studies in humans, but potential benefits may still warrant use of the drug. Category X is the lowest category, used when use

The parties agree that birth defects, including every type of birth defect alleged in this litigation, have occurred throughout history. For example, major congenital heart defects, which are among the most prevalent birth defects, occur in as many as 1% of live births. Expanding the scope to include all cardiac defects, one finds an incidence of approximately 7.5% of live births. Although some birth defects are caused by known genetic sources or environmental agents (such as certain viruses, radiation exposure, or teratogenic medications), most are due to currently unknown causes. Teratology is the scientific field which deals with the cause and prevention of birth defects.

Where plaintiffs allege that a medication, such as Zoloft, is a teratogen, it is common to put forth experts whose opinions are based upon epidemiological evidence. Although the “gold standard” for epidemiological studies is the double-blind, randomized control trial, such studies may not ethically be conducted on pregnant women. Therefore, in this context, epidemiologists must rely upon observational evidence.

Epidemiological studies examining the effects of medication taken during pregnancy on birth defects calculate a relative risk (RR) or odds ratio (OR).⁸ Simply speaking, these ratios are calculated by dividing the risk or odds of a particular birth defect in children born to medication users (exposed women)⁹ by the risk or odds of finding that birth defect in children born without prenatal exposure.¹⁰ Where the incidence of birth defects is approximately the same in

of the drug is not recommended for any pregnant women, as the risks clearly outweigh any benefits. One SSRI, Paxil, is a Category D drug, while all other SSRIs, including Zoloft, are Category C drugs.

⁸ Bérard Report at 9.

⁹ Studies differ in the methods used to identify whether women were exposed to a medication during pregnancy.

¹⁰ Other adverse outcomes may also be measured; for example, researchers might compare the number of miscarriages suffered by exposed versus unexposed women. The Court is not aware of any cases in the MDL in which the alleged injury is miscarriage.

medication-exposed and unexposed women, the RR or OR value will be close to one. The RR or OR is “interpreted as the increase in the risk of the outcome (congenital malformation) associated with the exposure of interest (SSRI) that is above and beyond the baseline risk.”¹¹

Researchers often statistically control for certain suspected and measurable confounding factors (*e.g.*, factors such as maternal age, weight, smoking, alcohol use, folic acid use, etc., which are correlated with exposure to the medication, and which may themselves contribute to an increased risk of the birth defect at issue), when information about those factors is available in the data set. When this is done, the researchers will report an “adjusted” ratio. The authors of the studies the Court has reviewed rely upon adjusted ratios, where available, when drawing conclusions.

Because an RR or OR calculation is only an estimate, the precision of which may be affected by general or study-specific factors (including confounders and biases, sample sizes, study methods, etc.), researchers also use statistical formulas to calculate a 95% confidence interval, which is an estimated range of plausible ratio values. A 95% confidence interval means that there is a 95% chance that the “true” ratio value falls within the confidence interval range. Some confidence intervals are narrow, indicating that the calculated rate ratio is fairly precise, and some are wide, indicating that it is not and that additional research is warranted. If the lower bound of the confidence interval is greater than one, researchers say that the ratio is “statistically significant” (*i.e.*, there is only a 5% chance that the increased risk reflected in the ratio is the result of chance alone), and will report finding a statistically significant *correlation* or

¹¹ Bérard Report at 9. The baseline risk varies by defect; however, the overall baseline population risk for major congenital malformations is approximately 3% in the United States.

association between the medication exposure and the birth defect at issue.¹² A statistically significant result does not necessarily indicate a large increase in risk; it simply indicates that the increased risk found is unlikely to result from chance alone.

Even where the confidence interval is narrow and the increased risk is statistically significant, teratologists will not draw firm conclusions from a single study, as apparent associations may reflect flaws in methodology, including multiple comparisons, bias, or confounding, or may be incongruous with existing scientific knowledge about biological mechanisms. When specific potential confounders or biases are identified, researchers will attempt to design studies in such a way that they can determine the degree to which those factors contributed to an outcome. In general, before concluding that there is a “true” association between a medication and an adverse outcome, the teratology community requires repeated, consistent, statistically significant human epidemiological findings, and studies which address suspected confounders and biases.¹³

Epidemiological studies alone can only inform scientists that two events (*e.g.*, medication use and a birth defect) are associated. For this litigation, the experts have been asked to opine as to whether Zoloft *causes* the birth defects at issue, which requires analysis beyond the identification of statistical correlations reported in published epidemiological studies. To infer a causal relationship from an association, scientists look at well-established factors sometimes referred to as the Bradford-Hill criteria. These include: the strength of the association between

¹² A factor may also be protective. For example, prenatal exposure to folic acid is associated with a decrease in neural tube defects. If a factor is protective, the ratio estimate will be less than one, and if the confidence interval’s upper bound is less than one, that protection is statistically significant.

¹³ “Absent consistent, repeated human epidemiological studies showing a statistically significant increased risk of particular birth defects associated with exposure to a specific agent, the community of teratologists does not conclude that the agent is a human teratogen.” *Wade-Greux v. Whitehall Labs., Inc.*, 874 F. Supp. 1441, 1453 (D.V.I. 1994), *aff’d* 46 F.3d 1120 (3d Cir. 1994).

the exposure and the outcome; the temporal relationship between the exposure and the outcome; the dose-response relationship; replication of findings; the biological plausibility of or mechanism for such an association; alternative explanations for the association; the specificity of the association (*i.e.*, does an outcome have only one cause, or several?); and the consistency with other scientific knowledge.¹⁴

III. DISCUSSION

Dr. Bérard has conducted epidemiological studies and published peer-reviewed papers examining the effect of maternal use of antidepressants during pregnancy. Although her opinions on the issues relevant to this litigation have evolved over time, Dr. Bérard's current opinions on the teratogenic effects of Zoloft are summarized in two paragraphs of the expert report she prepared for this litigation:

A. It is my opinion that SSRIs as a class of drugs cause an increased risk of adverse pregnancy outcomes, including spontaneous abortion and congenital malformations in multiple organ systems (including cardiac defects, craniosynostosis, pulmonary/respiratory defects, gastrointestinal defects (omphalocele, gastroschisis, pyloric stenosis, anal atresia), anencephaly, cleft lip and palate, neural tube defects, limb reduction defects, club foot, and PPHN [Persistent Pulmonary Hypertension of the Newborn].

B. It is my opinion that Zoloft causes an increased risk of spontaneous abortion and congenital malformations in multiple organ systems (including cardiac defects, craniosynostosis, pulmonary/respiratory defects, gastrointestinal defects (omphalocele, gastroschisis, pyloric stenosis, anal atresia), anencephaly, cleft lip and palate, neural tube defects, limb reduction defects, club foot, and PPHN.¹⁵

The Court must examine the reliability of the methods Dr. Bérard used to arrive at these opinions.

¹⁴ Bérard Report at 11-12.

¹⁵ Bérard Report at 6.

A. Methodological Issues Impacting Both Opinions

1. The Importance of Statistical Significance

As discussed above, in the field of epidemiology, the generally accepted method for determining whether a substance is a potential teratogen is to look for statistically significant associations between medication exposure and a pattern of birth defects, which are consistent and replicated across epidemiological studies, and to then apply the Bradford-Hill criteria. Dr. Bérard derives her conclusions about causation, in large part, by charting published findings from various studies (sometimes inaccurately) on a “forest plot” (a graphical depiction of the odds ratios and confidence intervals from multiple studies), and drawing conclusions from trends in odds ratios depicted on the forest plot without regard to whether the underlying published findings were statistically significant, and without further statistical analysis. Dr. Bérard testified that, in her view, statistical significance is certainly important within a study, but when drawing conclusions from multiple studies, it is acceptable scientific practice to look at trends across studies, even when the findings are not statistically significant.¹⁶ In support of this proposition, she cited a single source, a textbook by epidemiologist Kenneth Rothman, and testified to an “evolution of the thinking of the importance of statistical significance.”¹⁷

Epidemiology is not a novel form of scientific expertise.¹⁸ However, Dr. Bérard’s reliance on trends in non-statistically significant data to draw conclusions about teratogenicity, rather than on replicated statistically significant findings, is a novel methodology. Therefore, as the evidence on which Dr. Bérard relies is not evidence that derives from principles and

¹⁶ Bérard N.T. 4/9/14, 76:6-20.

¹⁷ Bérard N.T. 4/9/14, 76:21-77:13.

¹⁸ *In re Paoli*, 35 F.3d at 744 n.10.

techniques of “uncontroverted validity,”¹⁹ and hence is not readily admissible, the Court must determine whether Dr. Bérard’s opinion is admissible by applying the *Daubert* standard.

This is not a case where an expert is simply moving into novel terrain, wherein no methodology has yet been well established. There exists a well-established methodology used by scientists in her field of epidemiology, and Dr. Bérard herself has utilized it in her published, peer-reviewed work. The “evolution” in thinking about the importance of statistical significance Dr. Bérard refers to does not appear to have been adopted by other epidemiologists, even the very researchers she cites in her report. Her departure from that methodology in her litigation report and testimony requires more thorough justification than she has presented to the Court. Although she cited the Rothman textbook as support for her methods, she has advanced no evidence indicating that this is a “methodology [that] has been exposed to critical scientific scrutiny,”²⁰ or that it has been adopted by other scientists in the field.

The Court is particularly concerned about the risk of reaching an erroneous conclusion using Dr. Bérard’s methodology. Dr. Bérard opines that, although one cannot assume teratogenicity from one weak association in one study, one *can* assume teratogenicity based upon multiple weak associations found across many studies. However, an equally plausible conclusion from multiple studies finding only weak associations, not greater than one would expect by chance, is that the true association is weak; so weak that one cannot conclude that the risk is greater than that seen in the general population. This is, in fact, the conclusion most researchers in Dr. Bérard’s field have reached regarding the association between Zoloft and birth defects, even those cited by Dr. Bérard in support of her contrary opinion.

¹⁹ *Id.*

²⁰ *Wade-Greaux*, 874 F. Supp. at 1479.

The Court is mindful of its function as a gatekeeper; it is not for the courts to be the pioneers, forging new trails in scientific thinking, especially when that means departing from well-established research principles, such as the principle of statistical significance. The Court understands that it is difficult to measure small increases in risk when the risk is for a rare event. However, Dr. Bérard testified that the precision of an estimate *decreases* when an event is rare (as in the case of most birth defects);²¹ thus the danger of misinterpreting a given (imprecise) result as a true association, when it may be the result of chance alone, is greater, underscoring the importance of the concept of statistical significance in this context.

In *Wade-Greaux*, a well-respected pediatric, developmental, and genetic pathologist, Dr. Gilbert, “testified that she does not believe that repeated, consistent epidemiological studies showing a statistically significant increased risk of malformations associated with the use of a medicine are required to reach a scientifically valid conclusion as to whether that medication can cause malformations in humans at therapeutic dosage.”²² This is very similar to the position that Dr. Bérard takes in this case. The *Wade-Greaux* court noted that Dr. Gilbert’s conclusions were not derived from methodology generally accepted by the teratology community, and further noted that Dr. Gilbert’s own published research on teratology recognized and adopted the generally accepted approach. Because Dr. Gilbert’s opinion was not based upon repeated, consistent epidemiological studies showing statistically significant increased risks, the court excluded her opinion and testimony. Similarly, this Court finds that Dr. Bérard has failed to demonstrate that her reliance on non-statistically significant findings is accepted within her scientific community. Like the experts in *Wade-Greaux*, Dr. Bérard is only able to draw her

²¹ Bérard N.T. 4/9/14 at 77:5-11.

²² *Wade-Greaux*, 874 F. Supp. at 1455.

conclusions “by ignoring the basic requirements of the relevant scientific community’s methodology.”²³

2. Analyzing Data from Multiple Studies

Because birth defects are rare, Dr. Bérard testified that the scarcity of replicated, statistically significant findings may be the result of insufficient power in the studies, even in large, population based studies, including thousands of exposed women. Therefore, Dr. Bérard testified, her approach was to look for “trends” in ratio estimates reported in (selected) studies, and draw conclusions about teratology from those trends, from rather than statistically significant results. She demonstrated this method to the Court using a forest plot.

“[T]he party presenting the expert must show that the expert’s findings are based on sound science, and this will require some objective, independent validation of the expert’s methodology.”²⁴ When epidemiologists hypothesize that there is a “true” association which individual studies are underpowered to detect at a statistically significant level, the widely accepted approach to combining data from multiple studies—thus increasing the power to detect an association—is to conduct a systematic meta-analysis. Dr. Bérard did not address, in her report or her testimony, her reasons for relying upon her novel method of examining trends in odds ratios to test her hypothesis, rather than relying upon the well-established method generally relied upon by epidemiologists,²⁵ nor did she provide objective, independent validation for her novel method.

²³ *Id.* at 1478.

²⁴ *Daubert v. Merrell Dow Pharms., Inc.*, 43 F. 3d 1311, 1316 (9th Cir. 1995).

²⁵ One meta-analysis, conducted by Nicholas Myles and his colleagues, was recently peer-reviewed and published. Myles 2013. Dr. Bérard did not discuss this meta-analysis in her expert report, due to the timing of its publication, but she testified that she is familiar with it. In contrast to Dr. Bérard’s litigation method, the Myles study used a well-established method for analyzing data from multiple studies, a meta-analysis following the guidelines for Meta-Analysis of Observational Studies in Epidemiology. The Court notes that, according to the

B. Opinion Regarding Teratogenicity of SSRIs as a Class

Dr. Bérard's opinion relies, in large part, on her presumption that SSRIs, although distinct from one another in chemical structure and pharmacokinetic properties,²⁶ will have similar teratogenic effects. Plaintiffs argue that it is appropriate methodology to interpret research studies which report associations between SSRIs, collectively, and birth defects as evidence that Zoloft, particularly, would increase the risk of birth defects to the same degree. The Court disagrees. Even assuming sound evidence of a common biological mechanism by which all SSRIs could plausibly impact fetal development (e.g. evidence of serotonin and related developmental pathways from the embryonic and animal studies), such evidence would only give rise to the *hypothesis* of a class-wide effect among the SSRIs when used in human pregnancies; because even small differences in chemical structures can result in different teratogenic effects, it is necessary to test the hypothesis before concluding that all SSRIs are similar in their teratogenic properties.

Examination of the very peer-reviewed, published epidemiology literature cited by Dr. Bérard in her report yields little evidence of a class effect. If there were a class effect, one would expect to find consistent associations between each drug in the class and a given outcome.

Methods section of the published paper, Myles made an effort to be as inclusive as possible of high-quality published studies, not excluding any studies on the basis of their results or conclusions. This meta-analysis involved the statistical analysis of the included studies and measured the statistical significance of the results.

As Dr. Bérard pointed out in her testimony, Myles did exclude some studies from his meta-analysis. However, there is no evidence that he selected studies for inclusion and exclusion based upon the extent to which they supported his *a priori* hypothesis. In contrast, Dr. Bérard admitted that her expert report and forest plot focused upon published studies which were most supportive of her opinion.

The Court further notes that Myles concluded that Zoloft was *not* significantly associated with major or cardiac malformations. While Dr. Bérard's conclusions are not at issue here, the contrary conclusions Myles reached using a well-established method raises the possibility that Dr. Bérard's decision to rely on an alternative, non-statistical method of assessing data from multiple studies may be driven by her desire to confirm her *a priori* hypothesis that Zoloft is a teratogen, rather than by her desire to test the possibility that individual studies were underpowered to detect true associations.

²⁶ Kornum 2010 at 30.

Instead, there are only scattered statistically significant associations, both within and between studies. For example, in the Louik 2007 study, researchers found a statistically significant association between Zoloft (sertraline) and septal defects, and paroxetine and right ventricular outflow path obstruction, but no association between other commonly prescribed SSRIs (fluoxetine and citalopram) and either of those birth defects. Similarly, that study found only one of the studied drugs was significantly associated with neural tube defects (paroxetine), limb reduction defects (sertraline), and omphacele (sertraline). The Malm 2011 study found a statistically significant association between fluoxetine and cardiovascular anomalies, and citalopram and neural tube defects. Other studies which studied multiple birth defects and multiple SSRIs also found scattered, statistically significant associations between certain birth defects and one (or occasionally two) of the four or more SSRIs studied.²⁷

Dr. Bérard opines that the absence of evidence supporting her opinion that every SSRI is associated with a similar increase in birth defects can be explained by study samples which are too small to measure the “true” association between individual SSRIs and birth defects. But this assertion that class effects would be detectable given significantly large samples of pregnant women is itself only a hypothesis, and there is scant evidence in support of that hypothesis, despite large population-based studies using the Danish and Finnish birth registries.²⁸

At first glance one such study, the Jimenez-Solem 2012 paper, appears to support a limited class effect for cardiovascular defects.²⁹ That study found statistically significant associations

²⁷ See Alwan 2007, Kornum 2010, Pederson 2009, Diav-Citrin 2008, Malm 2011, Colvin 2011, and Baaker 2010.

²⁸ See Jimenez-Solem 2012, Kornum 2010, and Pederson 2009 (Danish data); and Malm 2011 (Finnish data).

²⁹ Other epidemiologists have noted that heart defects are heterogeneous, and a specific exposure is not expected to increase the risk for congenital heart defects in general; therefore, it is important for researchers to

between three SSRIs (sertraline, fluoxetine, and citalopram) and heart defects. However, that study also found similar rates of heart defects in SSRI users who “paused” SSRI use before and during pregnancy. This study suggests that the statistical association observed between SSRIs and heart defects may result from an unmeasured confounding factor, unrelated to the medications, which impacts fetal development. Dr. Bérard’s opinion about class effects does not adequately address this study and its implications. If she disagrees with the methods, findings, or the conclusions, she must provide a detailed, scientific critique which explains her disagreement.³⁰ She has failed to do so.

The Myles meta-analysis was designed with the goal of directly comparing the teratogenic potential of individual SSRIs, to determine whether any are comparatively safer. That study noted that “the teratogenic potential of specific agents might differ from the aggregate result for SSRI . . . medications as a class,”³¹ but the researchers’ *“a priori* hypothesis was that SSRI medications (fluoxetine, paroxetine, sertraline, and citalopram) are individually associated with an increased risk of major, minor, and cardiac malformation in those infants exposed *in utero*.³² The statistical analysis performed on the aggregate data, however, did not support the *a priori* hypothesis. Fluoxetine and paroxetine were both significantly associated with major malformations overall (including cardiac malformation); only paroxetine was significantly

perform subanalyses on specific heart defects. Bakker 2010 (studying exposure to Paxil and the incidence of septal defects and other specific heart defects).

³⁰ *In re Avandia Mktg., Sales Pracs. & Prods. Liab. Litig.*, No. 07-1871, 2011 WL 13576, at *9 (E.D. Pa. Jan. 4, 2011); *In re Rezulin Prod. Liab. Litig.*, 369 F. Supp. 2d 398, 425 (S.D.N.Y. 2005) (“if the relevant scientific literature contains evidence tending to refute the expert’s theory and the expert does not acknowledge or account for that evidence, the expert’s opinion is unreliable. Accordingly, courts have excluded expert testimony ‘where the expert selectively chose his support from the scientific landscape.’” (citing *Carnegie Mellon Univ. v. Hoffmann-LaRoche, Inc.*, 55 F.Supp.2d 1024, 1039 (N.D.Cal.1999)); *see also Miller v. Pfizer, Inc.*, 196 F.Supp.2d 1062, 1086-87 (D.Kan.2002) (examining studies reaching contrary conclusions is “appropriate to minimize the likelihood of a false conclusion”), *aff’d*, 356 F.3d 1326 (10th Cir.2004).

³¹ Myles 2013 at 2.

³² *Id.*

associated with cardiac malformations; and sertraline and citalopram were not significantly associated with either major or cardiac malformations. The study found that SSRIs were not significantly associated with minor malformations. The authors recommended that “future research be directed towards individual agents, rather than SSRIs as a class.”³³ Dr. Bérard explained that she did not consider this paper when writing her report because of the timing of its publication, but testified that she is familiar with it. She criticized the paper for excluding certain studies from the meta-analysis, but she did not identify which studies were excluded or why this undermines her confidence in the work, and in particular its conclusions about class effects.³⁴

The Court also notes that the FDA does not treat SSRIs as a class with regard to warnings about use during pregnancy. Paroxetine is in category D, while all other SSRIs are in category C.

In her own published, peer-reviewed research, Dr. Bérard’s has opined that paroxetine may have uniquely teratogenic properties among the SSRIs.³⁵ She first published this opinion in 2007, and has published that opinion as recently as 2012.³⁶ Dr. Bérard fails to explain how she reconciles her earlier research and conclusions with her current opinion that all SSRIs are similarly teratogenic. Although Dr. Bérard explained that her opinion has evolved since 2012, when she concluded that there was no indication to stop using any antidepressant, except paroxetine, during pregnancy, the Court does not find support for the evolution of her opinion in the published epidemiology research, suggesting that her opinion is not grounded in generally

³³ Myles 2013 at 9.

³⁴ Tr. 4/9/14 at 86:8-18.

³⁵ The Court makes no finding as to the actual or relative safety of paroxetine or the other SSRIs; the Court cites Dr. Bérard’s work only for the purpose of establishing that Dr. Bérard has studied the issue of class effects and has opined to her peers, and in other litigation, that the SSRIs are not all similarly teratogenic (*i.e.*, there is no class effect).

³⁶ Bérard, 2007; Simoncelli 2010 (Zoloft and other non-paroxetine SSRIs the “first-line therapy” for treatment of depression during pregnancy); Santos 2012. In addition to her published research, in 2009, she was retained by plaintiffs as an expert in the Paxil (paroxetine) litigation, and opined that current data did not suggest that SSRIs as a class were teratogenic. Bérard Paxil Report (Doc. No. 691, Ex. 8).

accepted scientific methods.³⁷ While sound epidemiological evidence of a class effect may develop in the future, if Dr. Bérard’s current hypothesis is correct, that hypothesis needs to be tested using sound scientific methods. As Dr. Bérard’s expressed opinion regarding class effects is not evidence based, and is directly contrary to the findings of her own peer-reviewed, published research, the Court holds that Dr. Bérard’s reliance on class wide data to draw inferences about the teratogenic impact of Zoloft is not based on a scientifically reliable methodology, and cannot be put before a jury.

Accordingly, under the principles set forth in *Daubert*, Dr. Bérard’s opinion that SSRIs as a class of drugs cause an increased risk of adverse pregnancy outcomes must be excluded pursuant to Rule 702.

C. Opinion Regarding Teratogenicity of Zoloft, Specifically

1. *Study Selection and the Epidemiological Evidence*

Pfizer argues that Dr. Bérard reaches her conclusions using flawed methods, including the “cherry-picking” of studies and of findings within studies which support her position.

The Court finds that the expert report prepared by Dr. Bérard does selectively discuss studies most supportive of her conclusions, as Dr. Bérard admitted in her deposition,³⁸ and fails to account adequately for contrary evidence, and that this methodology is not reliable or scientifically sound.³⁹ The Court is especially concerned that Dr. Bérard failed to discuss her

³⁷ To the extent that Dr. Bérard attributes her altered opinion to developments in the biological mechanisms literature (i.e., evidence from outside her field of expertise), see discussion at page 22, *infra*.

³⁸ Bérard Dep. at 260:13-23; 185:2-186:1; 197:17-22, 258:22-259:5, 308:23-309:6, 320:23-321:3.

³⁹ *In re Avandia*, 2011 WL 13576, at *9; *In re Bextra & Celebrex Mktg. Sales Pracs. & Prods. Liab. Litig.*, 524 F. Supp. 2d 1166, 1176 (N.D. Cal. 2007); *In re Rezulin Prod. Liab. Litig.*, 369 F. Supp. 2d 398, 425 & n.164 (S.D.N.Y. 2005).

own peer-reviewed, published studies, in which she concluded that paroxetine, but not the other SSRIs, increased the risk of major congenital and cardiac malformations.⁴⁰

At the hearing, Dr. Bérard provided her rationale for excluding certain studies from her report. For example, she testified that she excluded studies, including some of her own, which used an active comparison group (i.e. a group of women taking another type of antidepressant which is not a suspected teratogen).⁴¹ However, she herself has written that conducting studies with an active comparison group adds valuable information to the literature as a whole.⁴² As a causal conclusion requires examination of the literature as a whole, Dr. Bérard's explanation does not justify ignoring studies using active comparators.

Dr. Bérard also pointed out methodological flaws or weaknesses in other studies. The Court notes that all epidemiological studies will have some weaknesses; studies of potential teratogens are particular prone to biases and confounders, because it is unethical to utilize a randomized, double-blind study to examine possible teratogenic effects of a drug. It is not entirely clear that the weaknesses in the studies Dr. Bérard excludes are greater than those in the studies upon which she relies. Moreover, rather than simply ignoring certain studies, the accepted scientific practice is for an expert to explain why she gives more weight to certain studies in forming her opinion, discussing methodology, power, and other key factors.⁴³ Therefore, although Dr. Bérard testified that her opinion was based upon the totality of the evidence, and she was able to articulate a rationale for excluding certain studies during the

⁴⁰ Bérard 2007; Simoncelli 2010; Santos 2012.

⁴¹ Bérard N.T. at 81:24-82:16.

⁴² See Bérard, 2007, at 25 (“given that antidepressant users have very different characteristics than non-users with regard to lifestyle, history of co-morbidity, including mood and anxiety disorders, and socio-economic status, all of which can be difficult to measure with precision (Ramos et al, 2005), we do not believe that having a population-based comparator as was done by Kallen and Otterblad Olausson (2006) is optimal.”)

⁴³ *In re Avandia*, 2011 WL 13576, at *9.

hearing, the Court shares Pfizer’s concern that Dr. Bérard’s opinion relies upon a selected subset of evidence without sufficient analysis of contrary evidence—a significant methodological weakness.⁴⁴

“Cherry-picking” is always a concern, but is of heightened concern in this case, where many of Dr. Bérard’s conclusions and opinions were formulated by identifying trends in odds ratio estimates selected from the published literature, rather than being based upon replicated, statistically significant odds ratio estimates. The fact that her conclusions are drawn from trends she observed in a self-selected subset of supportive studies, not the totality of the epidemiological evidence, further underscores her problematic methodology.

Moreover, the studies Dr. Bérard does rely upon in her report do not adequately support her opinions, especially in light of her change in opinion from 2012 to the present. Only one of the studies she relies upon was published in 2012, and none were published later. Two of the studies she relies upon were published in 2007, when Dr. Bérard was firmly of the opinion (expressed both professionally and as an expert witness for plaintiffs in the Paxil litigation) that Paxil was uniquely teratogenic and other SSRIs should be used as first-line medications for treating depression in pregnancy.⁴⁵ These studies both cautioned that the statistically significant findings reported might be an artifact of multiple comparisons. Yet Dr. Bérard relies upon them as evidence of an increased risk of birth defects in the exposed group without addressing this issue. Three of the studies she relies upon used overlapping data from the Danish birth registry, and the last of those studies demonstrated that any increased risk might be due to a confounding

⁴⁴ In contrast, Pfizer’s expert, Dr. Srivastava, testified that in his expert report, he discussed the research papers that he thought were most informative regarding the issue at hand, not those he thought best supported his opinion. Tr. 4/14/14, 82:2-19.

⁴⁵ Louik 2007; Alwan 2007.

factor and not the SSRI medication at all.⁴⁶ The remaining studies did not find any statistically significant association between Zoloft use and congenital malformations.⁴⁷ Thus, although Dr. Bérard’s “cherry-picking” does concern the Court, the Court is also concerned by the lack of scientific support for Dr. Bérard’s altered opinion in the selected studies upon which she does rely. An opinion based on subjective belief, rather than grounded in science, is not admissible.

2. Unsupported Causal Inferences

When drawing causal inferences from associations between exposure to a drug and an adverse outcome, scientists consider certain well-established causation factors, the Bradford-Hill criteria. These criteria include the strength of the association between the exposure and the outcome; the temporal relationship between the exposure and the outcome; the dose-response relationship; replication of findings; the biological plausibility of or mechanism for such an association; alternative explanations for the association; the specificity of the association; and the consistency with other scientific knowledge. An expert need not consider or satisfy all criteria in order to support a causal inference.

Pfizer argues that the Bradford-Hill criteria should only be applied after an association is well established, and that there is no well-established association between Zoloft exposure during pregnancy and birth defects. However, because the Bradford-Hill criteria include as factors the strength of the association between exposure and outcome, and replication of findings, the Court will not adopt Pfizer’s view but rather will examine the evidence put forth by Dr. Bérard with regard to each Bradford-Hill criterion.

⁴⁶ Pederson 2009; Kornum 2010; and Jimenez-Solem 2012.

⁴⁷ Malm 2011; Colvin 2011; Merlob 2009; Baaker 2010. An additional study, on which Dr. Bérard is the senior author (Nakhai-Pour 2010), showed an association between SSRI exposure and spontaneous abortions. This effect was not specific to Zoloft, and moreover, Dr. Bérard failed to link an increased risk of spontaneous abortions to any of the birth defects at issue in this litigation.

As discussed above, despite her assertions to the contrary,⁴⁸ the strength of the associations between exposure to Zoloft and various birth defects at issue is weak, often not greater than one would expect by chance alone, and replication of statistically significant associations is rare. Even in the studies cited by Dr. Bérard in her report, statistically significant findings in one study are not replicated in other studies. A notable exception is the Louik 2007 study, which found that Zoloft is significantly associated with septal defects in the heart; this was replicated in the Pederson 2009 and Kornum 2010 studies using the Danish registry data. While this replication is noteworthy, Dr. Bérard's opinion is not limited to one injury or even one organ system; she opines that Zoloft causes an increased risk of congenital malformations in multiple organ systems, including cardiac defects, craniosynostosis, pulmonary/respiratory defects, gastrointestinal defects, anencephaly, cleft lip and palate, neural tube defects, limb reduction defects, and club foot. Yet only one study she cites finds Zoloft is significantly associated with limb reduction; that same study finds Zoloft is significantly associated with gastro-intestinal defects.⁴⁹ A second study finds that Zoloft is significantly associated with an increased risk of anacephaly. These findings are not replicated in other studies. Not a single study cited by Dr. Bérard finds a statistically significant increased risk of craniosynostosis, respiratory defects, cleft lip and palate, neural tube defects, or club foot in Zoloft exposed offspring. Thus, Dr. Bérard begins her causation analysis with very weak evidence of an association, except in the case of septal defects, yet she expresses "the same degree of scientific certainty for her opinion that

⁴⁸ "[There is] extensive peer-reviewed scientific evidence demonstrating a strong and replicated causal association between exposure to SSRIs in general and Zoloft specifically and congenital malformations of multiple organ systems . . ." Bérard Report at 30-31.

⁴⁹ Louik, 2007.

Zoloft specifically causes heart defects, neural tube defects, gastrointestinal defects, limb defects, craniosynostosis, cleft lip and palate, spontaneous abortions, and PPHN.”⁵⁰

Additionally, sound scientific methodology requires that a scientist consider all of the scientific evidence when making causation determinations. The evidence of an association between Zoloft and birth defects, when one looks at the entirety of the literature, and not just the studies cited in Dr. Bérard’s report, is even weaker. As Pfizer notes, “Dr. Bérard is only able to present an illusion of ‘consistency’ because she selectively cites only findings that (she says) support her opinions.”⁵¹ As discussed above, Dr. Bérard’s selective reliance on studies which support her opinion is inconsistent with a valid and scientific methodology.

The next consideration is the temporal relationship between the exposure and the outcome. Scientists agree that organogenesis occurs in the first trimester of pregnancy, and therefore most of the studies reviewed by Dr. Bérard focus on first trimester exposure to SSRIs. However, as discussed above, one study found a similarly increased risk of birth defects even when women stopped taking sertraline months before becoming pregnant.⁵² Dr. Bérard has not reconciled this finding, which shows the same risk in the absence of a temporal relationship, with her opinion that sertraline adversely impacts organogenesis through first trimester exposure. In other words, the Jimenez-Solem study raised significant issues with regard to temporality, as well as confounding factors and biological mechanisms, and Dr. Bérard does not address any of them. Another study, co-authored by Dr. Bérard, specifically examined whether the duration of

⁵⁰ Doc. No. 831, at 11.

⁵¹ Doc. No. 691-1, at 21.

⁵² Jimenez-Solem 2012. *See also* Ramos 2008.

antidepressant use during the first trimester of pregnancy was associated with the occurrence of birth defects, and found that it was not.⁵³

With regard to dosage effects, one study conducted by Dr. Bérard found that paroxetine was associated with an increased risk of major cardiac and major congenital malformations only when pregnant women were exposed to more than 25 mg/day during the first trimester; exposure was not associated with birth defects when lower doses of paroxetine were used.⁵⁴ However, that study found no association between other SSRIs and birth defects, regardless of dosage.⁵⁵ The 2012 study by Jimenez Solem found no dose-response association.

In drawing conclusions about causation, researchers must also consider the biological plausibility of the association. Dr. Bérard testified that she has reviewed the *in vitro* and *in vivo* animal studies on the impact of serotonin on fetal development, and concludes that they support her opinions about causation. The Court will not discuss the substance of those studies in detail here, as Plaintiffs' biological mechanism experts' methods and conclusions will be examined in a separate opinion. However, the Court notes that the biological mechanism research does not, at this time, establish: 1) that each of the three developmental pathways hypothesized to be impacted by serotonin exist in humans; 2) the ideal range of serotonin in the developing organism (of any species); or 3) the range of serotonin present in the developing embryo when a pregnant woman is exposed (or unexposed) to Zoloft in pregnancy. In addition to the many unanswered questions about the proposed mechanism, *in vitro* and *in vivo* animal studies are

⁵³ Ramos, 2008.

⁵⁴ Bérard 2007.

⁵⁵ Bérard 2007.

“unreliable predictors of causation in humans,” in the absence of consistent data from human epidemiologic studies.⁵⁶

Dr. Bérard must also consider alternative explanations for the associations seen in the studies she relies upon, especially in light of the lack of consistency and replication discussed above. Some of these associations may be statistical artifacts of multiple comparisons.⁵⁷ Two other issues mentioned repeatedly in the literature are detection bias and confounding by indication. Detection bias means that an abnormality, especially a mild or asymptomatic one, is more likely to be detected in the exposure group, often due to increased medical vigilance of exposed infants.⁵⁸ Many researchers mention this issue in their studies, some even run analyses to try to address it,⁵⁹ but Dr. Bérard did not adequately discuss the possibility that detection bias accounted for the associations upon which she relies for her opinion. With regard to confounding by indication, as discussed above, Jimenez-Solem compared the outcomes for those pregnant women who paused their use of Zoloft before and during pregnancy to those who continued its use. Both the exposed group and the paused group had similarly increased risks of having a child with congenital birth defects. This suggests that it may not be SSRI exposure which accounts for the increased risk of birth defects, but rather one or more confounding risk factors found in both past and present SSRI users.⁶⁰ In her report and testimony, Dr. Bérard selectively cited this

⁵⁶ *Wade-Greaux*, 874 F. Supp. at 1483; see also *Daubert*, 43 F.3d at 1320 (holding that experts relying on animal studies should “point to some authority for extrapolating human causation from teratogenicity in animals.”)

⁵⁷ See Louik 2007.

⁵⁸ This may occur because the parents seek more comprehensive prenatal and postnatal testing, or because infants born to SSRI users have an increased risk of neonatal distress (e.g., because they are pre-term or have SSRI withdrawal symptoms), and thus are more likely to require intensive medical care in the early days of their lives.

⁵⁹ Kornum 2010.

⁶⁰ See also Ramos 2008 (Dr. Bérard is a co-author), which found that the most common congenital malformation was an atrial septal defect, which was somewhat more common in children born to women who used any type of antidepressant (SSRI or non-SSRI), as well as women who stopped use of an antidepressant during the first trimester of pregnancy.

study's exposed group findings in support of her opinion, ignoring without explanation the findings for the paused group and their implications with regards to confounding by indication.

IV. CONCLUSION

Dr. Bérard opines that SSRIs, in general, and Zoloft, in particular, cause a wide range of birth defects when used during pregnancy. Other researchers in her field have concluded that the epidemiological research on which Dr. Bérard relies provides no conclusive evidence of an association between Zoloft and birth defects. Many go further and advocate use of Zoloft as a first-line drug for treating depression in pregnancy. This does not represent a mere professional difference of opinion; Dr. Bérard's opinions regarding Zoloft are only made possible by her departure from use of well-established epidemiological methods. Dr. Bérard's methodology involved a rejection of the importance of replicated statistically significant epidemiological findings demonstrating an association between Zoloft and a pattern of birth defects, substituting a novel technique of drawing conclusions by examining "trends" (often statistically non-significant) across selected studies. Her methods are not scientifically sound. Additionally, in her report, Dr. Bérard failed to acknowledge and distinguish or otherwise address the research findings contrary to her litigation opinion, including her own peer-reviewed, published research. In summary, Dr. Bérard takes a position in this litigation which is contrary to the opinion she has expressed to her peers in the past, relies upon research which her peers do not recognize as supportive of her litigation opinion, and uses principles and methods which are not recognized by the relevant scientific community and are not subject to scientific verification. Because the methodology and reasoning underlying Dr. Bérard's opinion is not scientifically valid, the Court holds that Dr. Bérard's opinion is not grounded in the methods and principles of science, and therefore it does not satisfy Rule 702, and must be excluded.